The Annals of Statistics
1989, Vol. 17, No. 4, 1479–1486

BOOTSTRAPPING EXPLOSIVE AUTOREGRESSIVE PROCESSES

BY I. V. BASAWA, A. K. MALLIK, W. P. MCCORMICK AND R. L. TAYLOR

University of Georgia

Asymptotic validity of the bootstrap is established for the least squares estimate of the parameter of an explosive first-order autoregressive process. It is noted that nonnormal limit distributions are obtained for both the traditional and the bootstrap estimates. The theoretical bootstrap validity results are supported by appropriate simulation.

1. Introduction. It is now well known that Efron's bootstrap method provides a useful tool for studying the distributional properties of various statistics of interest. See Efron and Tibshirani (1986) for a recent review. The bootstrap can be used as an alternative to the conventional sampling distributions of statistics for both finite samples and asymptotics. In a seminal paper, Bickel and Freedman (1981) established the asymptotic validity of the bootstrap. See also Singh (1981) for an important contribution to the rates of convergence of bootstrap estimates. In a series of papers, Beran (1982, 1984, 1986) discussed various aspects of asymptotics for bootstrap estimates, confidence sets and test statistics for both parametric and nonparametric problems. Freedman (1981, 1984) studied bootstrap asymptotics for regression, and stationary econometrics models.

The current literature on the bootstrap is predominantly concerned with independent observations. Freedman (1984), however, discusses a stationary model involving dependent observations. Freedman's (1984) model also covers implicitly the stationary autoregressive process with a nonzero intercept. Bose (1988) discusses Edgeworth correction by bootstrap methods in stationary autoregressive processes. Our main aim in this paper is to establish the asymptotic validity of the bootstrap estimate for a *nonstationary* (explosive) first-order autoregressive process $\{X_1, X_2, \cdots\}$ defined by the difference equation

$$X_j = \beta X_{j-1} + \varepsilon_j \quad \text{with } X_0 = 0, \tag{1.1}$$

where $\{\varepsilon_j\}$ is a sequence of independent and identically distributed random errors with an unknown distribution and $E(\varepsilon_j) = 0$, $\mathrm{Var}(\varepsilon_j) = \sigma^2$ (unknown). If $|\beta| < 1$, the process $\{X_j\}$ is (asymptotically) stationary and $|\beta| \geq 1$ corresponds to the nonstationary case. The nonstationary case is further divided into two cases: (a) explosive case, $|\beta| > 1$, and (b) unstable (or critical) case, $|\beta| = 1$. In this paper we discuss the explosive case; the unstable case will be considered

Received May 1987; revised February 1989.
AMS 1980 *subject classifications*. Primary 62M07, 62M09; secondary 62M10, 62E20.
Key words and phrases. Autoregressive process, bootstrap, asymptotics, nonstationary processes.

elsewhere. The least squares estimate of β is given by

$$\hat{\beta}_n = \left(\sum_{j=1}^{n} X_{j-1}^2 \right)^{-1} \left(\sum_{j=1}^{n} X_j X_{j-1} \right). \tag{1.2}$$

It is known that $\hat{\beta}_n$ is consistent for β for all values of β in $(-\infty, \infty)$; see Rubin (1950). The limit distribution of $\hat{\beta}_n$ is, however, drastically different for the three cases, viz., stationary, explosive and unstable. In general, the limit distribution of $\hat{\beta}_n$ is normal for the stationary case and nonnormal for the two nonstationary cases. The two nonstationary cases can be considered as examples of nonergodic models of Basawa and Koul (1979) [also, see Basawa and Prakasa Rao (1980), Basawa and Scott (1983) and Basawa and Brockwell (1984)].

In the next section, we summarize some preliminary results for the explosive case and describe the bootstrap estimate. The asymptotic validity of the bootstrap estimate is established in Section 3. Some simulation results are presented at the end of Section 3.

2. Some preliminary results and the bootstrap estimate. We first review some basic limit results for the explosive case. The details may be found in Anderson (1959). Throughout this section we assume $|\beta| > 1$ and consider the model defined in (1.1). Define

$$U_n = \sum_{j=1}^{n} \beta^{-(j-1)} \varepsilon_j \quad \text{and} \quad V_n = \sum_{j=1}^{n} \beta^{-(n-j)} \varepsilon_j, \tag{2.1}$$

where $\{\varepsilon_j\}$ are i.i.d. random variables with $E(\varepsilon_j) = 0$ and $\operatorname{Var}(\varepsilon_j) = \sigma^2$. It is easily seen that U_n and V_n are identically distributed for each n. If $\{\varepsilon_j\}$ are normal, U_n (and V_n) is also normal with mean 0 and variance $(1 - \beta^{-2n})(1 - \beta^{-2})^{-1}$. Anderson (1959) has shown that there exist random variables U and V such that

$$(U_n, V_n) \to_d (U, V) \quad \text{as } n \to \infty, \tag{2.2}$$

where U and V are independent and identically distributed. If $\{\varepsilon_j\}$ are normal, U and V are normal each with mean 0 and variance $(1 - \beta^{-2})^{-1}$. In the general case, U and V can be represented by

$$U \overset{d}{=} \sum_{j=1}^{\infty} \beta^{-(j-1)} \varepsilon_j \overset{d}{=} V,$$

where, in view of the assumption $E\varepsilon_j^2 = \sigma^2 < \infty$, the series is almost surely convergent by an easy application of the three-series theorem. We note the common characteristic function of U and V is

$$\phi(t) = \prod_{j=1}^{\infty} \phi_\varepsilon(\beta^{-(j-1)} t),$$

where $\phi_\varepsilon(\cdot)$ is the characteristic function of ε_j.

Let $\hat{\beta}_n$ denote the least squares estimate of β given by (1.2). The following theorem summarizes the limit distribution of $\hat{\beta}_n$ for the case $|\beta| > 1$. See Anderson (1959) for a proof.

THEOREM 2.1. *For $\hat{\beta}_n$ defined in* (1.2), *under the model* (1.1) *we have*, $|\beta| > 1$,

$$(\beta^2 - 1)^{-1}|\beta|^n(\hat{\beta}_n - \beta) \to_d V/U, \tag{2.3}$$

where U and V are defined in (2.2).

Note that, in particular, when $\{\varepsilon_j\}$ are normal the limit distribution (2.3) is a Cauchy distribution. In the general case, when $\{\varepsilon_j\}$ are nonnormal the limit distribution in (2.3) depends heavily on the distribution of ε_j and does not have a simple form. This makes it especially appealing to obtain a bootstrap approximation for the limit distribution in (2.3) when the density of $\{\varepsilon_j\}$ is unknown.

We now describe the bootstrap estimate. Let $\hat{\varepsilon}_j = X_j - \hat{\beta}_n X_{j-1}$ and define $\tilde{\varepsilon}_j = \hat{\varepsilon}_j - n^{-1}\sum_{j=1}^n \hat{\varepsilon}_j$, the centered residuals. Denote by $\tilde{F}_n$ the empirical distribution function based on $\{\tilde{\varepsilon}_j,\ j = 1, \dots, n\}$. Thus $\tilde{F}_n$ associates mass n^{-1} to each of $\tilde{\varepsilon}_j$, $j = 1, \dots, n$. Now, pretending that $\tilde{F}_n$ is the true distribution, draw a random sample $\{\varepsilon_j^*,\ j = 1, \dots, n\}$ from $\tilde{F}_n$. Thus, conditionally on $(X_1, \dots, X_n)$, the random variables $\{\varepsilon_j^*,\ j = 1, \dots, n\}$ are i.i.d. with distributed function $\tilde{F}_n$. Now, construct the bootstrap sample $\{X_j^*,\ j = 1, \dots, n\}$ recursively by

$$X_j^* = \hat{\beta}_n X_{j-1}^* + \varepsilon_j^*, \qquad j = 1, \dots, n, \tag{2.4}$$

with $X_0^* = 0$. The bootstrap least squares estimate is then given by

$$\hat{\beta}_n^* = \sum_{j=1}^n X_j^* X_{j-1}^* \Big/ \sum_{j=1}^n X_{j-1}^{*2}. \tag{2.5}$$

The main aim of this paper is to derive the conditional limit distribution of $\hat{\beta}_n^*$ given $(X_1, X_2, \dots)$ and to show that it is the same as the limit distribution given in Theorem 2.1.

We conclude this section with the statement of a technical result needed in the proof of the validity of the bootstrap procedure. We refer to Basawa, Mallik, McCormick and Taylor (1988) for a complete proof.

LEMMA 2.1. *For the model given in* (1.1) *with* $E\varepsilon_j = 0$, $\operatorname{Var}\varepsilon_j = \sigma^2 < \infty$ *and* $|\beta| > 1$, *we have*

$$\frac{1}{n}\sum_{j=1}^n (\tilde{\varepsilon}_j - \varepsilon_j)^2 \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \to \infty. \tag{2.6}$$

Moreover

$$\hat{\beta}_n \xrightarrow{\text{a.s.}} \beta \quad \text{as } n \to \infty. \tag{2.7}$$

3. Limit distribution of the bootstrap estimate. We can now state the main result of the paper.

THEOREM 3.1. *Conditionally on* $(X_1, X_2, \dots, X_n)$ *as* $n \to \infty$ *we have, for* $|\beta| > 1$,

$$(\hat{\beta}_n^2 - 1)^{-1} |\hat{\beta}_n|^n (\hat{\beta}_n^* - \hat{\beta}_n) \to_d V/U \tag{3.1}$$

for almost all sample paths $(X_1, X_2, \dots)$, *where* U *and* V *are defined in* (2.2), *and* $(\hat{\beta}_n, \beta_n^*)$ *are defined, respectively, in* (1.2) *and* (2.5).

PROOF. Define

$$U_n^* = \sum_{j=1}^{n} \hat{\beta}_n^{-(j-1)} \varepsilon_j^* \quad \text{and} \quad V_n^* = \sum_{j=1}^{n} \hat{\beta}_n^{-(n-j)} \varepsilon_j^*. \tag{3.2}$$

The result in the theorem can be deduced analogously to that of Theorem 2.1 provided we show that, conditionally on $(X_1, \dots, X_n)$, $(U_n^*, V_n^*) \to_d (U, V)$ for almost all sample paths, where U_n^*, V_n^*, U and V are as defined earlier. This is because the limit distributions of $\hat{\beta}_n^*$ and $\hat{\beta}_n$ are determined respectively by those of (U_n^*, V_n^*) and (U_n, V_n). In what follows we shall show that the conditional characteristic function of (U_n^*, V_n^*) given $(X_1, \dots, X_n)$ converges to the characteristic function of (U, V), for almost all sample paths $(X_1, X_2, \dots)$. Let $E^*(\ )$ denote the expectation with respect to the distribution $\tilde{F}_n$ (of ε_j^*) conditional on $(X_1, \dots, X_n)$. Let $\phi_{U_n^*}(t) = E^*(\exp itU_n^*)$.

It is shown first that

$$\phi_{U_n^*}(t) = \prod_{j=1}^{n} E^*\left(\exp it\hat{\beta}_n^{-(j-1)} \varepsilon_j^*\right) \tag{3.3}$$

converges a.s. to $\phi_U(t)$ as for each $t \in R$ where the null set does not depend on t. To accomplish this it suffices to show

$$\lim_{m \to \infty} \sup_{n \geq m} \sum_{j=m}^{n} \left| E^*\left(\exp it\hat{\beta}_n^{-(j-1)} \varepsilon_j^*\right) - 1 \right| = 0 \quad \text{a.s.}, \tag{3.4}$$

where the convergence in (3.4) is uniform on compact sets in R and

$$\lim_{n \to \infty} \left| E^*\left(\exp it\hat{\beta}_n^{-(j-1)} \varepsilon_j^*\right) - \phi_{\varepsilon_1}\left(t\beta^{-(j-1)}\right) \right| = 0 \quad \text{a.s.} \tag{3.5}$$

for each $j \geq 1$ and uniformly in t in bounded intervals. First,

$$\begin{aligned} \left| E^*\left(\exp it\beta_n^{-(j-1)} \varepsilon_j^*\right) - 1 \right| &\leq E^*\left(|t\hat{\beta}_n^{-(j-1)} \varepsilon_j^*|\right) \\ &= |t| \, |\hat{\beta}_n|^{-(j-1)} \frac{1}{n} \sum_{k=1}^{n} |\tilde{\varepsilon}_k|. \end{aligned}$$

Since by Lemma 2.1 $\hat{\beta}_n \overset{\text{a.s.}}{\to} \beta$, $|\beta| > 1$ and $(1/n)\sum_{k=1}^n |\tilde{\varepsilon}_k| \overset{\text{a.s.}}{\to} E|\varepsilon_1| < \infty$, it follows that for almost each sample path there exist positive integers C and m such that $(1/n)\sum_{k=1}^n |\tilde{\varepsilon}_k| \leq C$ and $|\hat{\beta}_n| \geq \delta > 1$ for all $n \geq m$. Thus, for almost each sample path,

$$\left|E^*\left(\exp it\hat{\beta}_n^{-(j-1)}\varepsilon_j^*\right) - 1\right| \le |t|\delta^{-(j-1)}C$$

for all $n \ge m$, and (3.4) is established. Next,

$$\begin{aligned}
&\left|E^*\left(\exp it\hat{\beta}_n^{-(j-1)}\varepsilon_j^*\right) - \phi_{\varepsilon_1}\left(t\beta^{-(j-1)}\right)\right| \\
&\quad \le \left|\frac{1}{n}\sum_{k=1}^{n} \exp it\hat{\beta}_n^{-(j-1)}\tilde{\varepsilon}_k - \frac{1}{n}\sum_{k=1}^{n} \exp it\hat{\beta}_n^{-(j-1)}\varepsilon_k\right| \\
&\qquad + \left|\frac{1}{n}\sum_{k=1}^{n} \exp it\hat{\beta}_n^{-(j-1)}\varepsilon_k - \frac{1}{n}\sum_{k=1}^{n} \exp it\beta^{-(j-1)}\varepsilon_k\right| \qquad (3.6)\\
&\qquad + \left|\frac{1}{n}\sum_{k=1}^{n} \exp it\beta^{-(j-1)}\varepsilon_k - E\left(\exp it\beta^{-(j-1)}\varepsilon_1\right)\right|.
\end{aligned}$$

For the first term in (3.6),

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{k=1}^{n} \exp it\hat{\beta}_n^{-(j-1)}\tilde{\varepsilon}_k - \frac{1}{n}\sum_{k=1}^{n} \exp it\hat{\beta}_n^{-(j-1)}\varepsilon_k\right| \\
&\quad \le \frac{1}{n}\sum_{k=1}^{n} \left|\exp it\hat{\beta}_n^{-(j-1)}\varepsilon_k\right| \left|\exp it\hat{\beta}_n^{-(j-1)}(\tilde{\varepsilon}_k - \varepsilon_k) - 1\right| \\
&\quad \le \frac{|t|}{n}\sum_{k=1}^{n} |\hat{\beta}_n^{-(j-1)}|\,|\tilde{\varepsilon}_k - \varepsilon_k| \\
&\quad \le \left(|t|\hat{\beta}_n^{-2(j-1)}\right)^{1/2}\left(\frac{|t|}{n}\sum_{k=1}^{n} (\tilde{\varepsilon}_k - \varepsilon_k)^2\right)^{1/2} \to 0 \quad \text{a.s.}
\end{aligned}$$

uniformly in t on compact subsets by Lemma 2.1. Similarly,

$$\left|\hat{\beta}_n^{-(j-1)} - \beta^{-(j-1)}\right|\frac{|t|}{n}\sum_{k=1}^{n} |\varepsilon_k| \to 0 \quad \text{a.s.}$$

uniformly in t on compact subsets, which implies that the second term in (3.6) goes to zero a.s. By the Glivenko–Cantelli theorem and the convergence theorem for characteristic functions, the third term in (3.6) goes to 0 a.s. and uniformly for t in a compact set. Hence, (3.5) is established, and it follows that

$$\phi_{U_n^*}(t) = \prod_{j=1}^{n} \phi_{\varepsilon_1^*}\left(t\hat{\beta}_n^{-(j-1)}\right) \overset{\text{a.s.}}{\to} \prod_{j=1}^{\infty} \phi_{\varepsilon_1}\left(t\beta^{-(j-1)}\right) = \phi_U(t). \tag{3.7}$$

The convergence in (3.7) is, for almost every sample path individually, uniform on compact subsets. Hence, one exception set can be used for all t. Similarly,

$$\phi_{V_n^*}(s) = \prod_{j=1}^{n} \phi_{\varepsilon_1^*}\big(s\hat{\beta}_n^{-(n-j)}\big) \to \sum_{j=1}^{n} \phi_{\varepsilon_1}\big(t\beta^{-(j-1)}\big) = \phi_U(s) = \phi_V(s).$$

For $m = [n/2]$,

$$\phi_{U_n^*, V_n^*}(t, s) = \sum_{j=1}^{m} \phi_{\varepsilon_1^*}\big(t\hat{\beta}_n^{-(j-1)} + s\hat{\beta}_n^{-(n-j)}\big) \prod_{l=1}^{n-m} \phi_{\varepsilon_1^*}\big(t\hat{\beta}_n^{-(n-l)} + s\hat{\beta}_n^{-(l-1)}\big).$$

Moreover,

$$\begin{aligned}
&\left|\prod_{j=1}^{m} \phi_{\varepsilon_1^*}\big(t\hat{\beta}_n^{-(j-1)} + s\hat{\beta}_n^{-(n-j)}\big) - \prod_{j=1}^{m} \phi_{\varepsilon_1^*}\big(t\hat{\beta}_n^{-(j-1)}\big)\right| \\
&\quad = \left| E^*\left(\exp i \sum_{j=1}^{m} \big(t\hat{\beta}_n^{-(j-1)} + s\hat{\beta}_n^{-(n-j)}\big)\varepsilon_1^*\right) - E^*\left(\exp i \sum_{j=1}^{m} t\hat{\beta}_n^{-(j-1)}\varepsilon_1^*\right)\right| \\
&\quad \le E^*\left[\left|\exp i \sum_{j=1}^{m} t\hat{\beta}_n^{-(j-1)}\varepsilon_1^*\right| \left|\exp i \sum_{j=1}^{m} s\hat{\beta}_n^{-(n-j)}\varepsilon_1^* - 1\right|\right] \\
&\quad \le \sum_{j=1}^{m} |s|\,|\hat{\beta}_n|^{-(n-j)} E^*|\varepsilon_1^*| \\
&\quad \le \left(\sum_{j=1}^{m} |s|\,|\hat{\beta}_n|^{-(n-j)}\right)\left(\frac{1}{n}\sum_{k=1}^{n} |\tilde{\varepsilon}_k|\right),
\end{aligned}$$

which can be made small a.s. by choosing n large since, by Lemma 2.1,

$$\frac{1}{n}\sum_{k=1}^{n} |\tilde{\varepsilon}_k| \xrightarrow{\text{a.s.}} E|\varepsilon_1| \quad \text{and} \quad |\tilde{\beta}_n| \xrightarrow{\text{a.s.}} |\beta| > 1.$$

Similarly,

$$\begin{aligned}
&\left|\sum_{l=1}^{n-m} \phi_{\varepsilon_1^*}\big(t\hat{\beta}_n^{-(n-l)} + s\hat{\beta}_n^{-(l-1)}\big) - \sum_{l=1}^{n-m} \phi_{\varepsilon_1^*}\big(s\hat{\beta}_n^{-(l-1)}\big)\right| \\
&\quad \le \sum_{l=1}^{n-m} |t|\big(|\hat{\beta}_n|^{-(n-l)}\big)\left(\frac{1}{n}\sum_{k=1}^{n} |\tilde{\varepsilon}_k|\right).
\end{aligned}$$

Thus, by virtue of (3.7) and the above, for each t and s,

$$\begin{aligned}
\phi_{U_n^*, V_n^*}(t, s) &\xrightarrow{\text{a.s.}} \prod_{j=1}^{\infty} \phi_{\varepsilon_1}\big(t\beta^{-(j-1)}\big) \prod_{j=1}^{\infty} \phi_{\varepsilon_1}\big(s\beta^{-(j-1)}\big) \\
&= \phi_U(t)\phi_V(s),
\end{aligned}$$

where the exceptional set can be chosen independently of t and s. Theorem 3.1 is thus proven. □

TABLE 1
Bootstrap cumulative distribution of T_n

$P(T_n \leq t)$	0.01	0.025	0.05	0.10	0.25	0.50	0.75	0.90	0.95
t	−2.356	−2.001	−1.652	−1.284	−0.661	−0.031	0.661	1.273	1.644
$N(0,1)$ (values)	−2.327	−1.960	−1.645	−1.282	−0.679	0.000	0.674	1.282	1.645

COROLLARY 3.1. *If $\{\varepsilon_j\}$ are assumed normal, we have, for $|\beta| > 1$, conditionally on $(X_1, X_2, \dots, X_n)$,*

$$T_n = \hat{\sigma}_n^{-1}\left(\sum_{j=1}^{n} X_{j-1}^{*2}\right)^{1/2}\left(\hat{\beta}_n^* - \hat{\beta}_n\right) \to_d N(0,1),$$

for almost all sample paths, where

$$\hat{\sigma}_n^2 = n^{-1}\sum_{j=1}^{n}\left(X_j - \hat{\beta}_n X_{j-1}\right)^2.$$

We omit the proof of this corollary for the sake of brevity. We refer to Basawa, Mallik, McCormick and Taylor (1988) for the details.

The cumulative distribution of T_n in Corollary 3.1 was simulated with $n = 200$, and $\beta = 1.05$ with $\{\varepsilon_j\}$ taken as $N(0,1)$ variates. Table 1 gives the simulation results using 5000 bootstrap samples. The $N(0,1)$ cumulative distribution values are given for comparison.

Acknowledgments. The first author's work was partially supported by a grant from the Office of Naval Research. We are indebted to Jack Reeves for his assistance in the simulation work reported above. Thanks are due to a referee and the Associate Editor for their constructive suggestions.

REFERENCES

ANDERSON, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30** 676–687.

BASAWA, I. V. and BROCKWELL, P. J. (1984). Asymptotic conditional inference for regular nonergodic models with an application to autoregressive processes. *Ann. Statist.* **12** 161–171.

BASAWA, I. V. and KOUL, H. L. (1979). Asymptotic tests of composite hypotheses for non-ergodic type stochastic processes. *Stochastic Process. Appl.* **9** 291–305.

BASAWA, I. V. and PRAKASA RAO, B. L. S. (1980). Asymptotic inference for stochastic processes. *Stochastic Process. Appl.* **10** 221–254.

BASAWA, I. V. and SCOTT, D. J. (1983). *Asymptotic Optimal Inference for Non-ergodic Models. Lecture Notes in Statist.* **17**. Springer, New York.

BASAWA, I. V., MALLIK, A. K., MCCORMICK, W. P. and TAYLOR, R. L. (1988). Bootstrapping non-stationary autoregressive processes. Technical Report, Dept. Statistics, Univ. Georgia.

BERAN, R. J. (1982). Estimated sampling distributions: The bootstrap and competitors. *Ann. Statist.* **10** 212–225.

BERAN, R. J. (1984). Bootstrap methods in statistics. *Jahresber. Deutsch. Math.-Verein.* **86** 14–30.

BERAN, R. J. (1986). Simulated power functions. *Ann. Statist.* **14** 151–173.

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

BOSE, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.* **16** 1709–1722.

EFRON, B. and TIBSHIRANI, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with discussion). *Statist. Sci.* **1** 54–77.

FREEDMAN, D. (1981). Bootstrapping regression models. *Ann Statist.* **9** 1218–1228.

FREEDMAN, D. (1984). On bootstrapping two-stage least squares estimates in stationary linear models. *Ann. Statist.* **12** 827–842.

RUBIN, H. (1950). Consistency of maximum likelihood estimates in the explosive case. In *Statistical Inference in Dynamic Economic Models* (T. C. Koopmans, ed.) 356–364. Wiley, New York.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

DEPARTMENT OF STATISTICS
UNIVERSITY OF GEORGIA
ATHENS, GEORGIA 30602